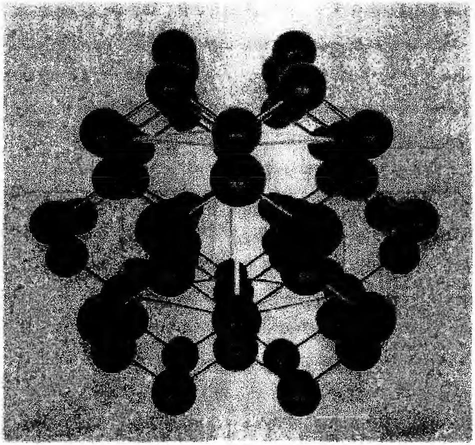
HANDBOOK OF SEMICONDUCTOR SILICON TECHNOLOGY



Edited by

William C. O'Mara, Robert B. Herring
and
Lee P. Hunt

NOYES PUBLICATIONS

Contents

4

Silicon Wafer Preparation

Richard L. Lane

1.0 INTRODUCTION

Integrated circuits and discrete solid state devices are manufactured on wafers made of silicon single crystal material, using a very complex series of processing steps. In order to obtain high yields and good device performance, it is important that the starting wafers be of reproducibly high quality because high resolution patterns are optically formed on the wafer, and the front surface must be smooth and flat on both a macro- and a microscale. The electrical and chemical properties of the wafer surface must also be well controlled and therefore the preparation of starting silicon wafers is a crucial portion of IC and device manufacture. This discussion will focus primarily on technical principles and theories that are applicable to silicon wafer preparation, and the practical application of those theories. Details, such as formulas, recipes, and specific process parameters are not given, because they vary considerably between different producers; likewise, specific equipment is not recommended, because the proper choice depends upon factors which are outside the scope of this discussion, such as plant capacity, labor costs, product specifications, and budget restrictions. For current equipment information, the reader is referred to the several directories (1-4).

Wafer preparation is the conversion of grown crystal

ingots into wafers which are ready for the wafer fab line. Each wafer must have at least one surface which is clean, flat and damage-free. In principle, a suitable wafer could be produced by cutting a crystal into thin slices and polishing one side of the slice until all of the saw marks are removed and the surface appears smooth and glossy. For some non-critical applications, such a wafer might be of sufficient quality. However, today's device and circuit designs are often more complex, and as circuit pattern geometries become smaller and wafer dimensions larger, the demands on wafer preparation become more stringent.

Additional preparation steps along with tighter process controls are required in order to achieve the required wafer quality.

Thus, silicon wafer preparation usually encompasses an extensive sequence of steps, starting with the as-grown crystal and ending with the polished wafer, ready for the fab line.

1.1 Wafer Preparation Processes

Figure 1 is a typical process flow diagram for wafer preparation. The three major categories, i.e. crystal shaping, wafer shaping, and wafer finishing are each represented by a series of process steps, some of which may be optional as indicated, depending upon the final wafer application. The right hand column includes some typical in-process control measurements that may be performed as the wafer preparation processes proceed.

These measurement techniques are discussed in the later section on in-process measurements. A brief statements of the purpose of each of the separate process steps follows below:

* Crystal cropping — Removal of the seed and tang ends of the crystal and any out-of-spec portions, using a circular diamond saw.

* Crystal grinding — The diameter of the crystal is reduced to the specified value and tolerance, usually with a diamond grinding wheel.

* Flat grinding — Flat areas are ground lengthwise along the crystal which serve later as identification of the wafer type and orientation.

* Crystal etching — The crystal is immersed in an etch bath which removes most of the surface getting from the grinding operation.

* Wafering — The crystal is cut into thin wafers, using a machine with an annular diamond blade (ID blade).

* Heat treatment — Crystals or sawn wafers are heat treated to eliminate oxygen donors, thereby normalizing resistivity.

* Edge Rounding — The as-sawn wafers are ground on the periphery with a diamond form wheel to remove the square corners and to produce the desired edge contour.

* Wafer Lapping — The thickness uniformity and flatness of the wafers are improved by the use of an abrasive slurry on a large flat lapping plate. Lapping may be either single or double sided.

* Wafer grinding — An alternative to lapping, using a diamond wheel instead of a lapping slurry. Commercial wafer grinders grind only one side of the wafer at a time.

* Wafer etching — Chemical etchants are used to remove damaged surface layers of the wafer from the sawing, lapping and edge profiling operations.

* Polishing — The matte surface of the lapped or sawn wafer is converted to a damage-free, specular surface. Polishing may be either single or double sided.

* Back side damage — A controlled amount of surface damage is applied to the back side of the wafer for gettering purposes.

* Cleaning — Contaminants in the form of particulates, organics, and inorganics are removed from the wafer surface by a series of chemical and mechanical cleaning operations, preparing the wafer for the device fab line.

* Marking — Wafers are marked individually for the purpose of identification and traceablility.

CATEGORY	PROCESS STEP	MEASUREMENT
	CHOPPING	OXYGEN CARBON RESISTIVITY
CRYSTAL SHAPING	O.D.GRINDING	DIAMETER
	FLAT GRINDING	FLAT WIDTH
	ETCHING	SURFACE ROUGHNESS THERMAL SLIP
	WAFERING	ORIENTATION THICKNESS FLATNESS, TAPER BOW, DAMAGE
WAFER SHAPING	* HEAT TREATMENT	RESISTIVITY
	EDGE CONTOURING	EDGE PROFILE
	* LAPPING	THICKNESS FLATNESS TAPER, DAMAGE
	WAFER ETCHING	SURFACE ROUGHNESS
	POLISHING	VISUAL INSPECTION
	CLEANING	THICKNESS FLATNESS VISUAL INSPECTION
WAFER FINISHING	* BACK SIDE DAMAGE	SURFACE ROUGHNESS
	* MARKING	
	CLEANING	VISUAL INSPECTION
	PACKAGING	

* Signifies Optional Process

Figure 1: Process Flow Diagram for Silicon Wafer Preparation

It should be noted that there is no "standard" sequence of steps for wafer preparation. The steps used in a typical operation will depend upon the use to which the wafers will be put and upon the specific procedures and quality control used in each successive step. For example, in some wafer preparation facilities polishing is performed directly on the as-sawn and etched wafers. In other operations, lapping is performed directly after wafering. Whether to use lapping depends upon (i) the final specification for the wafer (such as flatness) and (ii) the quality of the slicing operation.

Thus when lapping is performed after sawing, surface quality and thickness uniformity from the sawing operation may be relaxed somewhat, and the demands on the subsequent polishing process may also be less critical.

1.2 Silicon Removal Principles

The key processes listed above require the removal of silicon material to produce wafers of precise dimensions, and therefore wafer preparation is sometimes called "wafer shaping" in the industry. Material is removed from the crystal or the wafer by various removal processes, including mechanical, chemical, or a combination of mechanical and chemical means. Mechanical removal processes are sawing, lapping, and grinding using abrasives such as diamond, silicon carbide or aluminum oxide. These are frequently referred to as abrasive machining operations. Etching is a chemical removal process, whereas polishing combines both chemical and mechanical action to remove silicon.

According to Gielisse and Stanislao (5), material can be removed from a surface by one of two ways; (a) on a macroscale where the removed particles are much larger than atomic or molecular dimensions and (b) on a microscale in which the material is removed atom by atom or molecule by molecule. The distinction between macro- and microscale removal leads to the two fundamental types of removal processes: grinding and polishing. Macroscale removal requires penetration of the surface with an abrasive grit which produces localized forces in excess of the yield strength of the material. With grinding, the abrasive must be harder than the material being removed and the localized pressures are high in order to obtain penetration. Microscale removal leads to a very smooth surface and is called polishing. Polishing does not require such penetration and is by definition a low pressure process. The polishing agent is nearly always of lower hardness than the material being polished. Note: In this chapter the term "abrasive" is meant to be the particulate material which is used in a purely mechanical macroscale removal operation such as slicing, lapping or grinding. In the polishing operation, the term "polishing agent" is used rather than "abrasive" to distinguish between the two and to signify that polishing is more complex than a simple mechanical removal process. The differences between abrasive removal and removal by polishing are vividly evidenced by the roughness of the resulting surface, and the remaining lattice damage to be discussed later.

1.2.1 Mechanical Removal: At this point we limit the discussion to the abrasive machining of silicon. Unlike metals, plastics, and some amorphous materials, there is no plastic flow associated with abrasive machining of silicon. Silicon is a hard, brittle substance, and the penetration of abrasive particles establishes a field of damage in the form of cracks extending into the material from the point of penetration Figure 2. Intersection of the cracks leads to material removal by release of particles (6). Abrasive machining necessarily produces a rough surface and leaves sub-surface damage consisting of microcracks, dislocations and stresses. The magnitude of the roughness and the damage is directly related to the abrasive grit size. The speed of material removal is also related to grit size, (large grits remove material faster) and therefore as the wafer approaches final dimension, finer grit size may be used to more precisely control removal and to minimize remaining subsurface damage.

There are two modes of mechanical removal used in machining silicon: (i) bonded abrasive machining and (ii) free abrasive machining. Bonded abrasives are used on silicon for cropping, wafering, crystal and wafer grinding, and edge contouring. Virtually all bonded abrasive processes for silicon utilize diamond particles in a metallic or resin matrix. The diamond tool (wheel or blade) is fed into the workpiece (the silicon crystal) and the diamond abrasive grains which are held tightly in the matrix of the tool are forced to penetrate into the workpiece. The abrasive moves at high speed and intermittently impacts the workpiece as it moves through it. The appearance of a surface produced by bonded abrasives is usually a directional scratch pattern, depending upon the relative motions of the tool and the workpiece.

In free abrasive removal, a suspended abrasive in the form of a slurry is fed between the workpiece and a tool which is usually made of a hard material such as cast iron. Pressure applied to the tool forces the abrasive into the workpiece. Relative motion between the workpiece and the tool creates a rolling motion to the abrasive and a crushing action on the workpiece. The resulting surface has a uniformly matte character.

In silicon wafer preparation, it is conventional to use the term "grinding" when referring to the shaping of wafer surfaces with fixed abrasives, and "lapping" when indicating the free abrasive method, whereas in other industries, namely glass lens shaping, free abrasive removal is often called grinding.

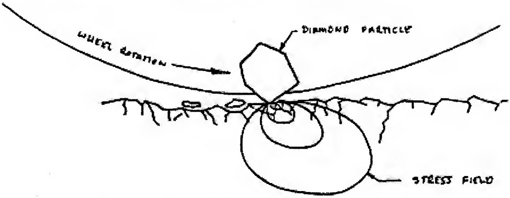


Figure 2: Abrasive Removal of Silicon

1.2.2 Chemical Removal: Etching is a chemical removal process whose principal application is to remove damaged surface material and the resulting stress caused by abrasive achining operations. The residual sub-surface damage in the form of dislocations and microcracks can propogate further into the bulk material if the wafer is placed under mechanical or thermal stress. Extensive propogation of cracks of course leads to fracture of the wafer, whereas extensive propogation of dislocations is called slip. Removal of the damaged layer by chemical means results in a stronger wafer which is resistant to chipping, breakage and slip from thermal stress.

Etching solutions either dissolve the silicon or react with it to form a soluble silicon compound. Etching is beneficial after any abrasive operation, however it is most common after crystal grinding, slicing, and lapping. Etching after lapping speeds the polishing process by removing material that would otherwise have to be removed by polishing. It also removes sawing or lapping damage from the back side of the wafer. If a wafer is polished without etching, the stress associated with the back side damage remains. When the stress is only removed on one side, the wafer may warp or bow.

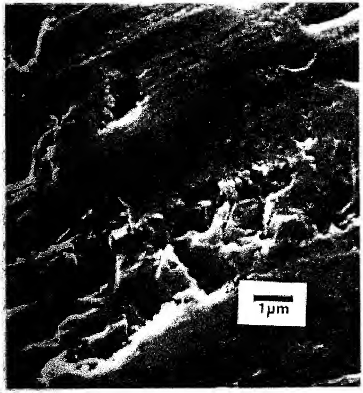
Etching rate is a sensitive function of temperature, concentration and agitation of the etch solution. As an etchant is used, its effectiveness diminishes. These variables combine to make etching a very difficult process to control. Most etchants tend to attack the silicon preferentially at the damaged locations where there is lattice strain, dislocations, or other discontinuities where the lattice is distorted and therefore more reactive, leaving surfaces that are not flat.

Some etchants, however are more selective than others. The use of an etchant of low selectivity produces a "chemically polished" surface with a shiny "orange peel" appearance. The SEM photomicrographs of Figure 3 illustrate such a surface as it is progressively etched. The as-sawn surface of the wafer is shown in Figure 3a in which the rough surface and oriented damage pattern from the diamond abrasive is evident. The wafer was step-etched for various times to remove damaged surface material to the depths indicated in Figure 3b. The removal of damage is indicated by the decrease in the number of etch pits as material is removed. The final picture after 48 micrometers of silicon have been removed is a typical chemically polished surface.

Etchants for silicon can be either acids or bases. Acid etchants are usually mixtures of various ratios of concentrated nitric, hydrofluoric, and acetic acid. The nitric acid initiates the reaction by forming a layer of silicon dioxide on the surface which is subsequently removed by the hydrofluoric acid. Acetic acid acts as a buffer to control the dissociation of the nitric acid. A common etchant for chemically polishing silicon is a mixture of 5 parts nitric, 1 part hydrofluoric, and 1 part acetic, by volume. When the wafers are immersed in this solution, usually in teflon carriers of 25 wafers, the reaction starts immediately, evolving a large amount of brown noxious, corrosive gases. As the reaction proceeds, the solution heats up, accelerating the etch rate. As more wafers are etched, the solution becomes depleted and the etching rate decreases. Eventually the acid is discarded and a new batch is prepared.

Disposal of both the evolved gases and the spent etchant is a serious problem. In a production facility, the evolved noxious gases must be removed from the work area but cannot be released into the atmosphere. Typically the fume hood used for acid etching must be connected to scrubbing and neutralizing equipment. Spent etchant is still highly acidic and requires complete neutralization before discharge into the sewer. Codes may vary in different localities, for example in the Silicon Valley area, the discharge of fluorides is restricted to 2-10 ppm., and therefore, simple neutralization will not be sufficient. In this case, fluoride must be precipitated to calcium fluoride, and the precipitated wastes hauled to an approved dump site (7). The proper handling of waste materials has become a major concern and proper adherance to federal, state and local codes is mandatory.

a



b

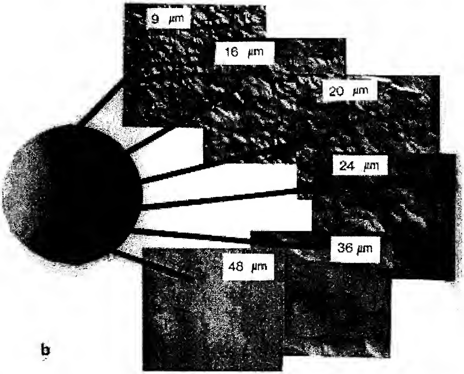


Figure 3: Photomicrographs of Etched and Unetched **Silicon**

The storage, handling and use of hydrofluoric acid also presents an important safety hazard because of the severe nature of the resulting burns. HF penetrates quickly and deeply into tissues, attacking the calcium and magnesium in bones, causing severe, slow-healing damage. If an HF burn is not quickly and properly treated, it can cause death.

Basic (or caustic) etchants have recently become more common in wafer preparation because of advantages in terms of cost, safety, and environmental problems. Potassium and sodium hydroxide are the most common caustic etchants, with potassium the preferred one. A 10 to 50% solution (by weight) of KOH in deionized water at 65-80 deg. C. removes silicon from the wafer surface at a reasonable rate. Because caustic etchants are more selective than acids, the resulting surface is not as smooth, however the damage and stress are removed as effectively as with acidic etching. Disposal of the spent caustic by-products is less of a problem than with acids.

1.2.3 Chemical-Mechanical Removal (Polishing): The polishing of silicon is in principle, similar to polishing other hard materials such as glass and ceramics. A polishing agent in a suspending fluid is applied to a soft pad such as felt or polyurethane and the workpiece is pressed against the rotating pad. A complete understanding of the polishing process has not been achieved, however it is generally agreed that it is not simply an extension of grinding or lapping with finer abrasives. There is ample evidence that polishing consists of interacting chemical and mechanical processes, and material removal rate is very slow compared to abrasive methods, because it is on an atomic or molecular scale. It is generally agreed however, that a substantial amount of material must be removed in order to obtain a satisfactory wafer surface. A discussion on polishing theory is presented below in the polishing section of this chapter.

2.0 CRYSTAL SHAPING

As-grown crystals have conical shaped ends. The seed end is quite short, perhaps with a height of one-fourth of the diameter, whereas the tang end of the crystal has a length approximately equal to the diameter. Automatic diameter control systems on crystal growth equipment are not capable of meeting the tight wafer diameter specifications that are required, so the crystals are usually grown one to

5.0 LAPPING

The slicing operation does not consistently produce the required flatness and parallelism for many wafer specifications. Since conventional polishing is not a flatness or thickness correcting process (see section 6, Polishing), lapping prior to polishing must often be used, because it is capable of achieving very precise thickness uniformity, flatness and parallelism. Lapping also prepares the surface for polishing by removing the sub-surface sawing damage, replacing it with a more uniform and smaller lapping damage.

5.1 Background

Lapping has long been used as a process to shape hard materials such as glass, stone, gems and ceramics, and to prepare them for the final polishing step. Even today the lapping technique is as much an art as it is a technology (32). The processes used for lapping and polishing silicon wafers evolved from the lens manufacturing industry which uses principles which were developed over several hundred years. A lapping wheel with the proper curvature is used as the tool to generate the desired lens shape. The machine holds the lens blank against the rotating lapping wheel and a silicon carbide slurry is fed onto the wheel while the lens is rotated and oscillated. If a fast removal rate is required, a coarse abrasive slurry is used, however the resulting surface will be rough and difficult to polish. Conversely, a fine abrasive removes material slowly, but the resulting surface is smoother and therefore the required polishing time will be shorter. Final lens polishing is performed on a similar machine, however the contoured wheel surface is covered with a polishing pad and the abrasive slurry is replaced with a polishing agent.

Silicon wafers are flat, and therefore the equipment is mechanically somewhat simpler than lens processing machines. During the last two decades, silicon wafer lapping techniques and materials have developed according to the needs of the semiconductor industry, and significant differences now exist between silicon and glass processing (33). For example, machines for lapping silicon wafers are large with many wafers lapped at a time. Wafers are thin, flexible and fragile compared to glass lenses, and therefore pressures, speeds, and abrasive materials are tailored to the material. Lens shaping usually requires single side processing because curva-

tures are often different on the two sides of the lens. Silicon wafers may be lapped by either a single or double side process. Double side processing is gaining in usage because improved wafer flatness can be achieved. Flatness of both the front and the back side is important since wafer fab lines use vacuum chucks to hold wafers for processing. With the trend toward larger diameter wafers and tighter dimensional tolerances, not only is double side lapping usually required but careful process control in the lapping step is necessary to obtain the required final wafer quality.

5.2 Current Technology

In its simplest form, a double side lapping machine consists of two very flat counter-rotating plates, carriers to hold and move the wafers between the plates, and a device to steadily feed abrasive slurry between the plates. The carriers are usually plastic templates of slightly smaller thickness than the wafers, with openings of a correct diameter to fit the wafers closely. The carriers are driven externally by gear rings which cause the wafers to move around the surface of the lapping plates in a planetary motion, thus the double side lapping machine is called a planetary lapper. A detailed description of the lapping process has recently been given by Dudley (34).

The lapping plates of the typical lapper are usually made of cast iron. Although other materials have been used, cast iron appears to have the right hardness for lapping silicon bacause the particles are not fully embedded in the lap, but move with a rolling, crushing action on the wafer to effect silicon removal. A softer lap would allow the abrasive to fully imbed in it which would cause the wafers to be scratched, whereas a hard lap would force the abrasive into the wafers, thereby causing the lapping plate to be worn rapidly, and the wafers to be damaged excessively.

Since the upper plate of the machine is floating, its axis of rotation adjusts to the upper surfaces of the wafers. Because of this, it is usually necessary to pre-sort the wafers before loading the machine, otherwise the parts may be tapered. If the sawing operation is closely controlled, sorting may not be required, except to reject the occasional out-of-spec wafer.

The wafers are manually loaded into the carriers, and they are flooded with slurry and the carriers are rotated in order to fully coat the wafers before lowering the upper

plate and starting the machine. This prevents breakage of wafers by providing an initial lubricity. The abrasive is typically a nine micrometer aluminum oxide grit. Aluminum oxide has a more blocky particle shape combined with a greater toughness than silicon carbide, resulting in a more controllable removal rate with less damage to the wafer. Commercial abrasives are suspended in water with proprietary additives to assist in suspension and dispersion of the particles, to improve the flow properties of the slurry, and to prevent corrosion of the lapping machine.

Lapping pressure of about 2 to 3 psi. (14 to 21 kpa.) is applied to the plates usually by hydraulic or air cylinders through most of the process, however starting pressure is kept low for 2 to 5 min. in order to allow the wafers to settle into the carriers, to allow the slurry to distribute uniformly, and to remove the high points of the wafers where the initial lapping forces are concentrated. Most lapping machines automatically ramp from the starting pressure up to the final pressure at a predetermined time in the lapping cycle.

The completion of lapping may be determined by elapsed time, or by a device on the machine which senses the actual wafer thickness during the process and stops the machine when the wafers have achieved that desired thickness. A sensing device has been reported (35) which is based upon piezoelectric material that is lapped simultaneously with the wafers. The material emits an electrical signal whose frequency is a function of the its thickness. Ultrasonic (36) and electronic (37) methods have also been reported. Mechanical methods to sense wafer thickness are usually not accurate enough to achieve the thickness tolerances required. Thickness tolerance achievable within a batch is +/- 0.0002 in. (+/- 0.005 mm.) for incoming wafers that are within +/- 0.001 in. (+/- 0.025 mm.).

Alhough lapping would appear to be simple in concept, the successful implementation of a production lapping operation requires the development of a technique and experience to achieve acceptable quality with good yields. Since flatness of the wafers is of primary importance, maintaining the flatness of the lapping plates is crucial. Small adjustments to the rotation rates of the plates and carriers will cause the plates to wear concave, convex or flat. Similarily, wafers can be made to be lapped convex or concave by adjusting rotation parameters (38). From time to time the plates will have to be brought back to a flat condition by the use of

special conditioning fixtures. As with the other processes used in wafer preparation, lapping will benefit by attention to optimizing process parameters, and the employment of trained and experienced personnel.

Commercially available equipment ranges over a wide size range, although all sizes are the same in principle. Machine manufacturers offer various features to improve productivity and yields and to reduce labor costs. The cost versus benefit tradeoff must be analyzed by the purchaser. Machine manufacturers also provide a base line process to the customer from which specific requirements can be developed. Figure 14 illustrates a commercial double side lapping machine.

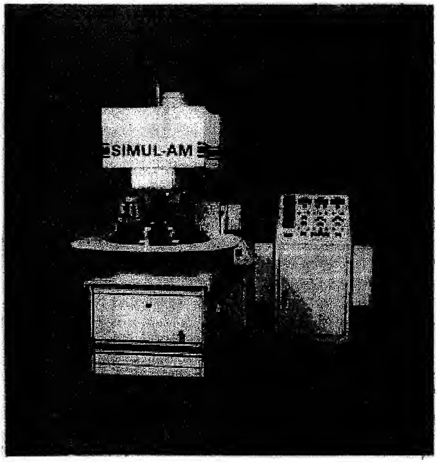


Figure 14: Commercial double side lapping machine. (Courtesy Kayex-Spitfire, a Unit of General Signal)

5.3 Wafer Grinding

Lapping is a messy process at best, and there has always been a desire to avoid it or to substitute an alternative operation for it. The most likely approach is wafer grinding. In grinding, the wafer is held on a vacuum chuck and a series of progressively finer diamond wheels is moved over the wafer while it is rotated on a turntable. The first wheel, a coarse one, removes the most material, and the subsequent wheels remove less material, produce less damage and create a smoother surface. Grinding leaves the wafers cleaner than lapping. It produces an oriented scratch pattern on the wafers which is noticeable to the eye when compared to lapping, however, there is reason to believe that the damage is no greater than that of a comparable sized lapping abrasive. Certainly there is little risk of imbedding particles of abrasive in the wafer. The most important drawback of wafer grinding is that only one side can be ground at a time. Because of this it may not be possible to obtain ground wafers that are as flat as lapped ones. Machines designed specifically for grinding silicon wafers are available in the market.

6.0 POLISHING

The purpose of polishing a silicon wafer is to produce a smooth surface that has properties which are as close as possible to those of the bulk material. The process must therefore take place uniformly across the surface, and it must not leave residual contamination or surface damage. Silicon wafer polishing derives from the glass lens polishing industry which, of course is very old. In principle, the same techniquies continue to be used, however some important modifications of the process have been developed to accommodate the special requirements of silicon wafer polishing.

6.1 Description of Polishing

If a lapped or ground silicon surface is examined with an electron microscope, it is found to contain a system of cracks, ridges and valleys closely resembling a rugged mountain terrain. Figure 15 schematically illustrates such an abraded surface. The peaks and valleys form a relief layer. Underlying the relief layer is a damaged layer characterized by microcr-

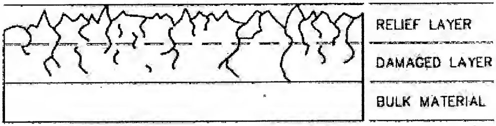


Figure 15: Cross sectional view of an abraded silicon surface.

acks, dislocations, slip, and stress. Both of these layers must be removed completely to expose the properties of the bulk material. When a much finer abrasive is used for lapping, the surface appearance is very much the same, only on a smaller scale. No matter how fine an abrasive is used, the lapped or ground surface will not achieve the smooth appearance of a polished one. This is the characteristic of brittle fracture of silicon which occurs in lapping and grinding. It indicates that the abrasive grains are moved across the surface under extremely high pressure, exceeding the elastic limit of the silicon.

Examinations of polished silicon or glass surfaces with the electron microscope have not shown any evidence of a relief surface such as that produced by lapping. Even at the highest resolution of the electron microscope which is about 15 angstroms, no scratches or ridges can be observed in a completely polished surface. It is therefore concluded that the polishing process is inherently different from the grinding process and that brittle fracture does not take place in polishing.

6.2 Historical Background

The bulk of the literature on polishing relates to the polishing of metals and glass. The polishing of metals is used principally to obtain the desired surface appearance as fast as possible and little concern is held for damage to the underlying bulk material. Metals are usually polished to produce a surface that is aesthetically pleasing, or to generate a certain roughness, (or smoothness), for example for the sealing surfaces in vacuum systems. Polishing of metals is considered to be simply an extension of fine grinding (39).

Glass and silicon polishing are quite different. Although

optical polishing experts might dispute it, it is believed that the surface quality required for silicon wafers exceeds that required for high quality optical glass, at least on a microscopic scale. In glass polishing, surface geometry is the principal objective, with damage and cleanliness of lower priority. Silicon wafers must be completely damage and haze-free, whereas geometry is not as critical. Typical wafer flatness is in the 1 to 2 micrometer range.

Although the objectives are different, the polishing of silicon resembles that of glass in many aspects, and therefore it is appropriate to review three theories that have been proposed for explaining the polishing mechanisms for glass:

Glass Polishing Theories

1. Glass polishing is purely a physical phenomenon, with the particles of polishing agent mechanically wearing away the material, much like grinding but on a smaller scale.

2. Polishing takes place because the surface exhibits a degree of plasticity. It can flow or be displaced when subjected to pressure, and in this way, a smoothing of the surface can take place. The plasticity is a result of thermal softening due to pressure and friction, or swelling due to the attack of water solutions which may produce a silica jel layer on the surface.

3. Polishing is possible because a great number of chemical and mechanical interactions occur. The chemical/mechanical characteristics of the polishing agent, the pad and the glass are affected by the suspending water solution. Low pressure contact results in atomic bonding between the polishing agent and the surface which removes material on a nearly atomic scale.

Herschel, (40) proposed the first theory by writing in 1830 that polishing was simply an extension of grinding. Beilby in 1903 (41) put forth the idea that polishing took place by flow of material from the peaks into the valleys due to thermoplastic deformation. He observed grooves in a

seemingly structureless polished surface after etching with hydrofluoric acid and concluded that polishing does not require removal of material, but occurs as a leveling process. Early observations by Lord Rayleigh in 1901 (42) were the first important ones which led to the conclusion that the polishing agent particles did not simply act as tiny cutters as Herschel had suggested. Klemm and Smekal, (43) demonstrated conclusively that a steel needle moved across a glass surface caused surface flow, and that this flow could fill in previously made grooves, however, there was not proof that such high pressures occur in the normal glass polishing process. A large portion of the literature on glass polishing deals with the question of whether glass flow is a significant effect in polishing. The work of French (44) is an example. French described a "beta" layer which is a distorted surface layer which can reach a thickness of about four micrometers. The beta layer corresponds to Beilby's flowed layer. The investigations of Bruche and Poppa, (45) and Kaller, (46) have indicated that the glass polishing phenomenon is probably explained by a combination of all three of these theories, each operating to a greater or lesser degree.

It is clear, in any case, that the removal of material in significant amounts is required to achieve high quality polishing. Just how this removal takes place is still not well understood. Conditions of extreme pressure and speed of travel have been required to demonstrate observable flow. It does not seem likely that plastic flow can be an important part of normal glass polishing, especially thermoplastic flow. In the case of polishing silicon, surface flow is even more improbable because of the crystalline nature of the material and its inability to soften like glass. Nevertheless, the current literature continues to refer to the "Bielby Layer" on a polished surface as a disordered or damaged layer resulting directly from the polishing process. The ideal polishing process for silicon should remove material without creating additional damage and without leaving a residual contaminated or disturbed layer. The current belief is that the optimum process is one in which a monomolecular layer of reaction product is produced on the surface and the polishing materials (pad, slurry, fluid) are able to remove the reaction material but must not abrade or otherwise damage the underlying material. Current literature on silicon polishing lends support to this theory (47).

Like glass surfaces, etching with a selective etchant such as Sirtl etch (48) will reveal subsurface damage (often

appearing as scratches) which are otherwise invisible. However, if the stock removal is sufficient to remove the original lapping damage, such a test only determines the quality of the polishing process (49). A high quality polished silicon wafer will be completely free of scratches or other polishing--related marks after etching.

6.3 Current Polishing Practice

Silicon wafers can be polished readily by the same pads and polishing agents that are used to polish glass, such as wool felt with iron oxide, (rouge) cerium oxide, or zirconium oxide (50). These agents, having a particle size in the 1 to 10 micrometer range, polish fast and produce a fairly good surface. Control of the pH of the slurry to a high value (around 9 to 10) gives the fastest removal rate, however the materials often produce a light haze on the surface that is difficult to remove.

Virtually all silicon wafers today are polished using a colloidal dispersion of silica in water (silica sol). The colloidal silica is an amorphous, synthetic sol produced by a gaseous reaction or by precipitation from a liquid, resulting in a sub-micron particulate material which is mixed with a high pH aqueous suspending medium. Typical particle size is in the range of 4 to 100 millimicrons (51). Additions of proprietary suspending agents are usually added to prevent settling of the particles. Slurries are shipped in concentrated form, and are diluted prior to use. The concentration of solids for polishing, (3 to 4%) is typically lower than with the fast glass polishing agents (52), and the removal rate achieved is lower, but the quality of the surface exceeds that of any other particulate polishing agent. In order to achieve a reasonable removal rate and still obtain the highest quality surface, polishing with colloidal silica is performed in two steps; stock removal and haze removal. As the names imply, the former is carried out with a higher concentration slurry and may proceed for about 30 minutes at a removal rate of about one micrometer per minute. The latter is done with a very dilute slurry, a softer pad, and for a short time, about 5 to 10 minutes. The amount of material removed is very small, usually less than one micrometer. Because of the active chemical reaction with colloidal silica agents, the wafers must be rinsed in deionized water immediately after polishing is complete to prevent haze or stains from forming again.

6.3.1 Polishing Variables: There are many variables which influence the rate and quality of polishing. Many of these are obvious. Increased pressure increases the removal rate, but excessive pressure leads to non-uniform removal, excessive heat generation, and fast pad wear. Removal rate is increased with high polishing temperatures, but also leads to haze formation due to the high chemical activity. High wheel speeds (expressed in feet per minute of travel of the wafers across the pad) increase removal rate, but also can lead to excessive temperatures, as well as problems in maintaining a uniform flow of slurry across the pad. High slurry concentration increases polishing rate, but is more costly and the resulting higher chemical/mechanical activity may result in haze formation. As the pH of the slurry increases, the removal rate increases gradually, until at a pH of about 12, the polishing rate falls off dramatically (49),(53). Further, at this critical pH the surface of the wafer becomes hydrophilic rather than the characteristic hydrophobic surface obtained at lower pH values.

Flatness: The commercial polishing process for silicon wafers is not a flatness improving process. At best it will not degrade the wafer flatness achieved in the lapping operation. A typical polishing pad may be 1000 micrometers thick, and is quite resilient. Pressure from the polishing machine compresses the pad as much as 250 micrometers. The flatness of the average wafer is in the one to two micrometer range. The pad therefore contacts the complete wafer surface at all times with nearly the same pressure, so that the removal rate is essentially uniform across the wafer. Since the total amount removed is 25 to 30 micrometers, an optimum polishing process is therefore one that maintains both chemical and mechanical uniformity of polishing conditions across the wafer. This is achieved by careful attention to all phases of polishing, including mounting of the pad, the slurry flow and concentration, alignment and flatness of the polishing plates, temperature control, wafer mounting procedure and the like.

Polishing Pads: Pad materials used for silicon are usually poromeric synthetic materials such as polyurethane. Many types are available with different thickness, hardness and surface roughness. The choice of the best pad material is a function of the chosen process variables. A very soft pad may have a tendency to produce an "orange peel" surface where a hard pad may have a tendency to glaze, remove material slower, and scratch the wafer. One usually attempts to use the hardest pad possible in order to obtain long pad

life, but soft enough to have a reasonably consistent removal rate.

Wafer Mounting: Wafers must be held in the polishing process in such a way as to firmly hold them against the polishing pad without bending, breaking, edge chipping or other damage. The wafer support must be firm because they are thin and flexible. There are two basic methods of mounting wafers for polishing; wax and waxless.

In wax mounting the wafers are attached to a flat carrier plate using a thermoplastic mounting wax. The plate is heated and the wax is coated on the plate. Great care must be exercised in assuring that the wax layer is of uniform thickness to fully support the wafers. Problems in achieving a uniform adhesive wax layer include entrapped air bubbles, foreign particles, and uneven coating of the wafer carrier plate. Bubbles and particles provide a source of localized pressure which tends to bow the wafer during the polishing process. When the wafer is released from the carrier plate there will be small regions where the wafer surface is not flat and the wafer thickness will vary in those regions. The patent by Walsh (54) describes a typical method and apparatus for wax mounting silicon wafers on a carrier plate.

After polishing, the wafers are removed from the carrier plate by heating until the wax is melted and carefully lifting them with tweezers. After removal, the wax and any polishing debris must be cleaned off. Needless to say, the wax mount/-demount process is time consuming and costly. Wafer breakage is often a problem because of the required handling. Thus the industry has sought for and attempted a number of "waxless" mounting techniques, some of which are in production.

The wafer mounting method must not only support the wafer uniformly against the polishing pressure, it must also resist the strong lateral force from the moving polishing pad. Wax mounting, of course accomplishes both of these objectives because the wax is an adhesive. Polishing with a waxless mounting method, sometimes called "free polishing" has to utilize some other means of resisting the lateral forces. Templates alone have been found to be unacceptable because the wafer is not capable of sustaining the high edge pressure without fracture or edge chipping. The usual method is to support the wafer on a somewhat resilient surface, such as a thin polyurethane pad which has sufficient frictional characteristics with the silicon to prevent the wafer from sliding. A template is usually added to assist in positioning the wafers

on the mounting plate and to prevent the occasional tendency of wafers to move.

Double side polishing, also necessarily a waxless form of polishing, is performed on a machine essentially like a double--sided lapper. There is considerable interest in double sided processing (both lapping and polishing) because of the potential for improved flatness, thickness and bow. However, double sided polishing has not as yet, received wide acceptance. Double side polishing, since it is performed with counter-rotating pads, tends to equalize lateral polishing forces on the wafers, thereby reducing the probability of edge chipping due to forces from the template.

Pressure: Although processing cost can be reduced by the faster removal rate achievable with higher pressure, there are tradeoffs. Higher pressure can lead to non-uniform removal, increased pad wear, poor temperature control, and breakage.

Temperature: Because of the chemical nature of polishing, an increase of temperature results in increased polishing rate. The optimum temperature which is typical in most polishing facilities is about 40°C. This is usually measured with an optical pyrometer sighting directly on the pad. Higher temperature can cause excessive evaporation of the slurry resulting in non-uniform removal rate. It can also lead to the formation of haze because of the faster chemical reaction, which presents difficulty downstream in the cleaning process.

6.3.2 The Optimum Polishing Process: Every wafer polishing operation has its own formula for polishing including polishing agent, pad, wafer holding method, polishing temperature, slurry flow rate, pressure, speed, and which has been developed on a particular machine to achieve the desired polished wafer quality. For this reason it is very difficult to describe a standard polishing technique. As the demand for flatter wafers increases, all wafer preparation processes become critcal, especially the lapping and polishing processes. The optimum polishing process for a given facility depends largely upon the interplay of product specification, yields, cost, and quality considerations, and must be developed uniquely for that facility.

6.3.3 Other Methods Of Polishing: Virtually all wafer production utilizes conventional slurry polishing, yet there is a need for flatter surfaces that are exceptionally clean and damage-free. There are many reported polishing processes but most consist of only minor variances from the general process described above. Three novel processes deserve

mention because they depart from the typical processes yet offer the possibility of improvements such as surface quality, speed, or flatness.

Copper ion exchange: A copper ion-containing solution is used in place of the polishing slurry (55). When the solution contacts the silicon wafer, copper ions are reduced to metallic copper. By the action of the polishing pad, the copper is removed along with the oxidized silicon, on a nearly atomic scale. The reported result is a very fast, damage-free polishing process. Although it is being used in production polishing, it has not met wide acceptance, perhaps because of the corrosive nature of the chemicals used, the cost, or concern about the possibility of copper contamination of the wafers.

Float polishing: A recently reported process (56), float polishing closely resembles pitch polishing where the pitch lap is replaced with a diamond-turned tin lap surface. The polishing agent is suspended in a liquid, and by controlling rotational speeds of the lap and the sample holder, the sample wafer is suspended on a film of the slurry. The polishing agent particles must be softer than the silicon, and removal is apparently effected by the bombarding of the silicon by the particles. Reportedly this method, although quite slow, yielded damage-free and scratch-free surfaces of exceptional flatness. "Non-contact" polishing, reported by Magee and Leung (57), appears to be a similar process to float polishing.

Mechano-chemical polishing: This method (58), also performed in a matter similar to conventional polishing processes is unique in that it is typically performed dry. The polishing agent must be a softer material than silicon, for example, calcium carbonate, and removal is apparently brought about by chemical reaction between the polishing agent and the wafer. Flat, scratch and damage-free surfaces have been observed using this method. As with float polishing, the process does not appear to have gained acceptance in production, yet the resulting superior surface may be beneficial for selected applications.

7.0 CLEANING

The topic of wafer cleaning is a complex one, and a comprehensive treatment of it would deserve a complete chapter in itself. There are several steps in the wafer

preparation sequence where cleaning may be appropriate, however this discussion includes only post-lap and post-polish cleaning since they are the most important cleaning processes in wafer preparation. For example, a cleaning procedure after slicing to remove mounting epoxy and swarf from wafers is not discussed, although this would be required in a typical facility. Slicing, lapping and polishing are relatively dirty operations compared to downstream wafer fabrication, and the cleaning processes used must be capable of effectively removing the relatively large amounts of particulate materials which are left on the wafers.

In addition to the removal of particles, a post-polish cleaning method should be able to remove organic materials such as oils, waxes and fingerprints, ionic subtances such as sodium and potassium, and trace metals. All of these contaminants are attached to the surface with bonding forces which may be physical, electrostatic or chemical in nature. Those materials that are chemically bonded (ions, atoms, films) are removed by chemical means whereas those that are less strongly attached (particles) may be removed by mechanical methods. The final cleaning step for polished wafers includes both scrubbing and chemical immersions followed by extensive rinsing in deionized water.

7.1 Mechanical Cleaning

Lapping residue consists of abrasive particles, iron particles from the lapping plates, debris from the wear of the carriers plus other contamination from the lapping machine area. Since the lapped wafers are sent to the polishing area after cleaning and etching, it is most important that all particluate contamination is removed to prevent scratching. Simple spray rinsing is not sufficient to assure the removal of all particles of lapping abrasive, however the use of high pressure (2000-3000 psi.) jets is effective in some cases. Scrubbing with high speed revolving brushes is very effective in removing particles. The brushes are made of hydrophilic fibers so that it is believed that the fibers do not actually contact the wafer surface, but ride on a thin film of attached water, providing the severe turbulence required to dislodge particles. Sonic "scrubbing" by the use of high frequency vibration in a water bath causes cavitation, and the collapse of the resulting bubbles effectively removes particles. Before polishing, thin films of ionic or organic contamination which may be on the wafers are of little consequence because the

polishing process will adequately remove these from the critical front surface.

In the case of single side processing with wax mounting, the slices are mounted on support plates before lapping, and remain mounted through the polishing step. In this case, the whole mounting plate with wafers attached must be cleaned free of all abrasive and silicon particles, since it passes through to the polishing area.

7.2 Chemical Cleaning

The chemical process that virtually all immersion techniques are based upon is the so-called "RCA Cleaning Method" described by Kern and Puotinen (59). This method utilizes a series of solutions designed to remove any type of contaminating material that is likely to be on the wafers. A summary of the procedure is as follows:

1. Molecular contaminants, such as gross organics, waxes, oils, etc. are removed with organic solvents such as freon with ultrasonic agitation followed by a freon vapor degrease.

2. Ionic contamination and trace organics are removed by immersion in a warm hydrogen peroxide-ammonium hydroxide solution, followed by a dilute hydrofluoric acid dip to remove the oxide film which is formed from the basic peroxide solution.

3. Atomic contamination such as adsorbed heavy metals are removed with a hydrogen peroxide--hydrochloric acid mixture.

Thorough rinsing in deionized water is required after the ionic and atomic removal, and of course the cleaning and rinsing processes must be performed in a clean room. Rinse water must be of the highest quality, otherwise contamination will be redeposited on the wafers. This means that the water must be free of dissolved and particulate contamination including bacteria (60),(61).

7.3 Other Cleaning Methods

Megasonic cleaning (62), is a recent development which appears to clean wafers as effectively as brush scrubbing and

chemical cleaning combined. It is a high frequency ultrasonic method using the RCA chemicals at room temperature with the wafers immersed as usual in carriers. Particles down to 0.5 micron are removed from wafers. It is reported to have greater throughput and significantly less chemical waste than chemical immersion cleaning.

Cleaning solutions based upon choline have been reported by Muroka, et. al. (63). This chemical which is similar to ammonia but is a fully ionized strong base, is combined with non-ionic surfactants to form an effective solution for removal of metal contaminants as well as bacteria. It may be used in place of or in addition to the RCA system. It is reported to be more useful in front end applications rather than for initial wafer preparation.

7.4 Equipment

The equipment used for cleaning can be as simple as a manually operated series of baths through which the wafer carriers are sequentially moved, to a fully automatic machine which eliminates the operator entirely. The cleanliness of the wafers will depend upon the ability of the system being used to first remove contamination and second to prevent redeposition, especially of particles. Since the human operator is the primary source of particulates in a clean room, the fully automatic system has a distinct advantage. Commercial equipment which is offered has been described recently in some detail, (64),(65). A wide variety of wafer scrubbers is available. For chemical cleaning, centrifugal spray units offer improvements over immersion baths since each wafer is uniformly exposed to uncontaminated, fresh chemicals.

The importance of high quality cleaning can not be overstressed as it is ultimately a yield determining factor. No cleaning system will be effective that is not protected from contamination from outside sources such as humans, bacteria, and miscellaneous particles as well as from impure chemicals and poor water. Since many contaminants are organic in nature, the use of strong oxidants is effective. Recently the direct and continuous application of ozone to deionized water systems has been reported (66) which significantly reduces bacteria in water, a leading cause of wafer contamination. High quality water is absolutely essential in wafer cleaning operations (67).